during the GID experiment did not reflect the sixfold symmetry of the COF lattice. This finding suggests that matching the COF lattice size and symmetry to the underlying graphene is not necessary to obtain crystalline films (see below).

The crystallinity and alignment of COF films on transparent SLG/SiO_2 substrates provides a means to organize functional π-electron systems within optoelectronic devices. Accordingly, films of two of the first COF semiconductors were grown on SLG/SiO_2. One of these frameworks, known as triphenylene-pyrene (TP)–COF (*11*), arises from incorporating a pyrene-2,7-diboronic acid linker in place of PBBA into the hexagonal COF-5 lattice (Fig. 4A). We obtained TP-COF in both thin-film and powder form using similar conditions to those described above (see figs. S14 and S15 for powder characterization and figs. S16 and S17 for top-down and cross-sectional micrographs). The GID of the films (Fig. 4B) indicates similar vertical alignment of the 2D lattice, as judged by the attenuation of the signals with increasing $Q_\perp$ and the absence of the out-of-plane 001 diffraction. The increased pore size of TP-COF is apparent from the prominent 100 diffraction at 0.19 $Å^{-1}$, and the 110 (0.34 $Å^{-1}$), 200 (0.39 $Å^{-1}$), and 210 (0.52 $Å^{-1}$) are also observed. Refinement of these data provided lattice parameters $a = b = 37.7$ Å in excellent agreement with those derived from PXRD data of TP-COF powders (37.5 Å) (*11*). The transparent SLG/SiO_2 substrate enabled ultraviolet/visible/near infrared (UV/Vis/NIR) spectroscopy of a COF film in transmission mode for the first time (Fig. 4C, red trace). The spectrum is consistent with the presence of both HHTP and pyrene chromophores and shows improved vibrational resolution of the absorbance bands relative to the diffuse reflectance spectrum of the powder sample. The photoluminescence of the film (Fig. 4C, blue trace) is characteristic of pyrene excimer emission over all excitation wavelengths, arising from efficient energy transfer from HHTP to pyrene that was observed in TP-COF powders.

Finally, we confirmed that COFs lacking hexagonal symmetry may also be crystallized on SLG by preparing a Ni phthalocyanine-PBBA COF on SLG/SiO_2 (Fig. 4D). GID of the film (Fig. 4E) again exhibited diffraction peaks localized near $Q_\perp = 0$ located at 0.27 $Å^{-1}$ (100), 0.55 $Å^{-1}$ (200), 0.81 $Å^{-1}$ (300), and 1.08 $Å^{-1}$ (400). These data correspond to a vertically aligned 2D square lattice with parameters $a = b = 23.0$ Å that match those obtained from the characterization of the powder sample. Cross-sectional images indicate a continuous film of ~210 ± 25–nm thickness (figs. S18 and S19). The translucent, turquoise films absorb strongly over the visible range of the spectrum as a consequence of the Ni phthalocyanine chromophores (Fig. 4F). Both the films and the powders are nonemissive, as is expected for H-aggregated phthalocyanines (see figs. S20 and S21 for powder characterization). These vertically aligned, porous phthalocyanine COFs are intriguing precursors of ordered heterojunction films long thought to be ideal for organic photovoltaic performance (*28*).

References and Notes

1. A. P. Côté *et al.*, *Science* **310**, 1166 (2005).
2. A. P. Côté, H. M. El-Kaderi, H. Furukawa, J. R. Hunt, O. M. Yaghi, *J. Am. Chem. Soc.* **129**, 12914 (2007).
3. R. W. Tilford, W. R. Gemmill, H. C. zur Loye, J. J. Lavigne, *Chem. Mater.* **18**, 5296 (2006).
4. D. F. Perepichka, F. Rosei, *Science* **323**, 216 (2009).
5. H. M. El-Kaderi *et al.*, *Science* **316**, 268 (2007).
6. L. Grill *et al.*, *Nat. Nanotechnol.* **2**, 687 (2007).
7. N. A. A. Zwaneveld *et al.*, *J. Am. Chem. Soc.* **130**, 6678 (2008).
8. A. Gourdon, *Angew. Chem. Int. Ed.* **47**, 6950 (2008).
9. M. In't Veld, P. Iavicoli, S. Haq, D. B. Amabilino, R. Raval, *Chem. Commun.* **2008** 1536 (2008).
10. S. Wan, J. Guo, J. Kim, H. Ihee, D. Jiang, *Angew. Chem. Int. Ed.* **48**, 5439 (2009).
11. S. Wan, J. Guo, J. Kim, H. Ihee, D. Jiang, *Angew. Chem. Int. Ed.* **47**, 8826 (2008).
12. E. L. Spitler, W. R. Dichtel, *Nat. Chem.* **2**, 672 (2010).
13. X. Ding *et al.*, *Angew. Chem. Int. Ed.* **50**, 1289 (2011).
14. X. Li *et al.*, *Nano Lett.* **9**, 4359 (2009).
15. A. Kumar, C. Zhou, *ACS Nano* **4**, 11 (2010).
16. A. Reina *et al.*, *Nano Lett.* **9**, 30 (2009).
17. K. S. Kim *et al.*, *Nature* **457**, 706 (2009).
18. X. Li *et al.*, *Science* **324**, 1312 (2009); 10.1126/science.1171245.
19. M. P. Levendorf, C. S. Ruiz-Vargas, S. Garg, J. Park, *Nano Lett.* **9**, 4479 (2009).
20. S. Bae *et al.*, *Nat. Nanotechnol.* **5**, 574 (2010).
21. Materials and methods are available as supporting material on *Science* Online.
22. D. M. Smilgies, D. R. Blasini, *J. Appl. Cryst.* **40**, 716 (2007).
23. B. Lukose, A. Kuc, T. Heine, *Chemistry* **17**, 2388 (2011).
24. J. L. Baker *et al.*, *Langmuir* **26**, 9146 (2010).
25. D. M. Smilgies, *J. Appl. Cryst.* **42**, 1030 (2009).
26. C. Berger *et al.*, *Science* **312**, 1191 (2006); 10.1126/science.1125925.
27. K. V. Emtsev *et al.*, *Nat. Mater.* **8**, 203 (2009).
28. A. C. Mayer, S. R. Scully, B. E. Hardin, M. W. Rowell, M. D. McGehee, *Mater. Today* **10**, 28 (2007).

Acknowledgments: This research was supported by startup funds provided by Cornell University and the NSF-funded CCI-I Center for Molecular Interfacing (CHE-0847926). This work is based on research conducted at the Cornell High Energy Synchrotron Source (CHESS), which is supported by the NSF and the NIH/National Institute of General Medical Sciences under NSF award DMR-0936384. We also made use of the Cornell Center for Materials Research facilities with support from the NSF Materials Research Science and Engineering Centers program (DMR-0520404). J.W.C. thanks the NSF for the award of a Graduate Research Fellowship; E.L.S. thanks the NSF for the award of the American Competitiveness in Chemistry postdoctoral fellowship (CHE-0936988); and M.G.S. (FA9550-07-1-0332) and J.P. (FA9550-09-1-0691) thank the Air Force Office of Scientific Research. We thank D. Smilgies for helpful discussions and K. Cox for technical illustration assistance. A provisional patent application based on this work has been filed by Cornell University (USA 61/382,093).

Supporting Online Material

www.sciencemag.org/cgi/content/full/332/6026/228/DC1
Materials and Methods
SOM Text
Figs. S1 to S21

11 January 2011; accepted 28 February 2011
10.1126/science.1202747

A Virophage at the Origin of Large DNA Transposons

Matthias G. Fischer[1] and Curtis A. Suttle[1,2,3]*

DNA transposons are mobile genetic elements that have shaped the genomes of eukaryotes for millions of years, yet their origins remain obscure. We discovered a virophage that, on the basis of genetic homology, likely represents an evolutionary link between double-stranded DNA viruses and *Maverick/Polinton* eukaryotic DNA transposons. The Mavirus virophage parasitizes the giant *Cafeteria roenbergensis* virus and encodes 20 predicted proteins, including a retroviral integrase and a protein-primed DNA polymerase B. On the basis of our data, we conclude that *Maverick/Polinton* transposons may have originated from ancient relatives of Mavirus, and thereby influenced the evolution of eukaryotic genomes, although we cannot rule out alternative evolutionary scenarios.

Transposable elements (TEs) affect the genetic landscape of eukaryotes by creating new alleles, influencing gene regulation, and rearranging chromosome structure (*1*). TEs can be divided into retrotransposons and DNA transposons. DNA transposons are further categorized into cut-and-paste TEs, rolling-circle TEs, and the self-synthesizing *Maverick* or *Polinton* (MP) TEs (*2*). MP TEs are 9 to 22 kilobase pairs (kb) long, encode up to 20 proteins, and are found in a wide range of eukaryotes, including protists, fungi, and animals (*3*–*6*). Their coding capacity varies among species, yet they contain a conserved set of genes (*4*, *5*). Most MP TEs encode a retroviral integrase, a viral protein-primed DNA polymerase B, an adenosine triphosphatase (ATPase) similar to the FtsK-HerA genome-packaging ATPases of double-stranded DNA (dsDNA) viruses, and a cysteine protease with homologs in adenoviruses. Another protein termed

[1]Department of Microbiology and Immunology, 1365–2350 Health Sciences Mall, University of British Columbia, Vancouver V6T 1Z3, Canada. [2]Department of Botany, 3529–6270 University Boulevard, University of British Columbia, Vancouver V6T 1Z4, Canada. [3]Department of Earth and Ocean Sciences, 6339 Stores Road, University of British Columbia, Vancouver V6T 1Z4, Canada.

*To whom correspondence should be addressed. E-mail: csuttle@eos.ubc.ca

PY, which is frequently encountered in MP TEs (*4*), shows structural similarity to the major capsid protein of phycodnaviruses (*7*). Because MP TEs genetically resemble viruses of the bacteriophage PRD1-adenovirus lineage (*8*), a viral origin of these transposons has been proposed (*5*, *7*), although a retroviral integrase in combination with the PRD1-adenoviral genes, as found in MP TEs, is not known in viruses.

The giant *Cafeteria roenbergensis* virus (CroV), with a 730-kb genome (*9*), replicates in the marine phagotrophic flagellate *C. roenbergensis*. We have discovered a much smaller virus associated with CroV infection (fig. S1) (*10*), which we named Mavirus (for Maverick virus). CroV production and host cell lysis were reduced by Mavirus infection (table S1), and Mavirus was not able to replicate in the absence of CroV, as demonstrated by quantitative polymerase chain reaction (qPCR) (table S2). We therefore classify Mavirus as a virophage sensu La Scola *et al*. (*11*), who coined the term "virophage" in reference to Sputnik, a true virus parasite of Mimivirus. According to this definition, virophages are predicted to have no nuclear phase in their infection cycle, to replicate in the virion factory of their host virus, and to be dependent on enzymes provided by their host virus, rather than by the host cell (*11*, *12*). The discovery of Mavirus suggests that virophages may be a common phenomenon. The Mavirus genome (Fig. 1A) is a 19,063–base pair (bp) circular dsDNA molecule with a G+C content of 30.26% and 20 predicted protein-coding sequences (CDSs) of an average length of 883 nucleotides (nt), resulting in a genome-coding density of 92.6%. A nearly perfect inverted repeat of ~50 bp was situated in the intergenic region between CDSs MV20 and MV01. This region has the potential to adopt a hairpin structure with a 43-bp stem and a 15-nt loop (Fig. 1B). The adenine at the top of the loop was designated position 1 of the genome. Analysis of the intergenic regions revealed a conserved promoter motif (fig. S2) preceding all 20 CDSs that is highly similar to a putative late promoter motif in CroV (*9*). This implies that Mavirus gene expression is governed by the transcription machinery of CroV during the late stage of infection. Ten Mavirus CDSs showed sequence similarity to proteins from eukaryotes, bacteria, retroviruses, and dsDNA viruses, including at least four genes found in Sputnik (table S3 and fig. S3). Mavirus and Sputnik have homologous genes encoding a capsid protein (fig. S4), a predicted DNA-pumping ATPase (fig. S5), a predicted cysteine protease (fig. S6), and a predicted GIY-YIG endonuclease/Zn-ribbon protein (fig. S3); however, the integrases and promoter motifs (*13*) of the virophages are unrelated. Indeed, about 80% of the CDSs appear unrelated, suggesting that any evolutionary connection between these viruses must be ancient.

In contrast, a closer evolutionary relationship exists between Mavirus and MP TEs, which share seven homologous CDSs (Fig. 1A). One of the most conserved genes in MP TEs is the rve-superfamily retroviral integrase (rve-INT; Pfam entry PF00665, 1e-16) (*3*–*5*), which contains a conserved CHROMO domain (PF00385, 5e-04) of ~60 amino acids capable of interacting with chromatin (*14*, *15*). Usually, rve-INTs are expressed as part of the POL polyprotein, but in eukaryotes they can occur independently of retroelements in the form of Ginger DNA transposons (*16*) and cellular integrases (c-integrases) (*15*). MP TE-encoded c-integrases are the closest relatives to Mavirus protein MV02 (fig. S7). Another highly conserved gene found in MP TEs and Mavirus (MV03) is the predicted protein-primed DNA polymerase B (PolB) (fig. S8). For both rve-INT and PolB sequences, the closest identified relatives to Mavirus were found in a genomic contig of the slime mold *Polysphondylium pallidum*, which contains a MP-like element that is partially syntenic to Mavirus (Fig. 1C). This putative MP TE is truncated, and, consequently, no terminal-inverted repeats or target-site duplications were identified. A search for Mavirus-like promoter signals in *P. pallidum* was also unsuccessful. An additional Mavirus CDS with homologs in MP TEs is MV15, which bears Walker A, Walker B, and P9/A32-specific motifs that define the DNA-pumping ATPases of the FtsK-HerA clade found in membrane-containing DNA viruses and MP TEs (fig. S5) (*17*). The adjacent CDS, MV16, encodes a predicted adenoviral cysteine protease with its closest homologs in Sputnik, MP TEs, and giant viruses (fig. S6). Homologs to three additional Mavirus genes—the superfamily 3 helicase, the major capsid protein, and MV19, which is similar to a cell wall surface anchor family protein from *Bdellovibrio* bacteriovorus—are occasionally encoded by MP TEs (*4*, *5*, *7*). Furthermore, both Mavirus and MP TEs have similar genome lengths (19 kb and 15 to 20 kb, respectively) and genome structures. MP TEs have terminal-inverted repeats of several hundred nucleotides in length and highly conserved ends that start with 5′-AG and end with CT-3′ (*4*). Similarly, cutting the circular genome of Mavirus between nucleotides 19,063 and 1 (Fig. 1B) would result in a linear DNA molecule with 5′-AG and CT-3′ termini, as well as 50-bp terminal-inverted repeats.

The observed parallels in genome length, genome structure, and gene content to MP TEs in general, as well as the sequence similarity and partial synteny to the MP-like element in *P. pallidum* in particular, imply that Mavirus and MP TEs have evolved from a common ancestor. The genetic similarities of these transposons and dsDNA viruses have spawned various evolution-

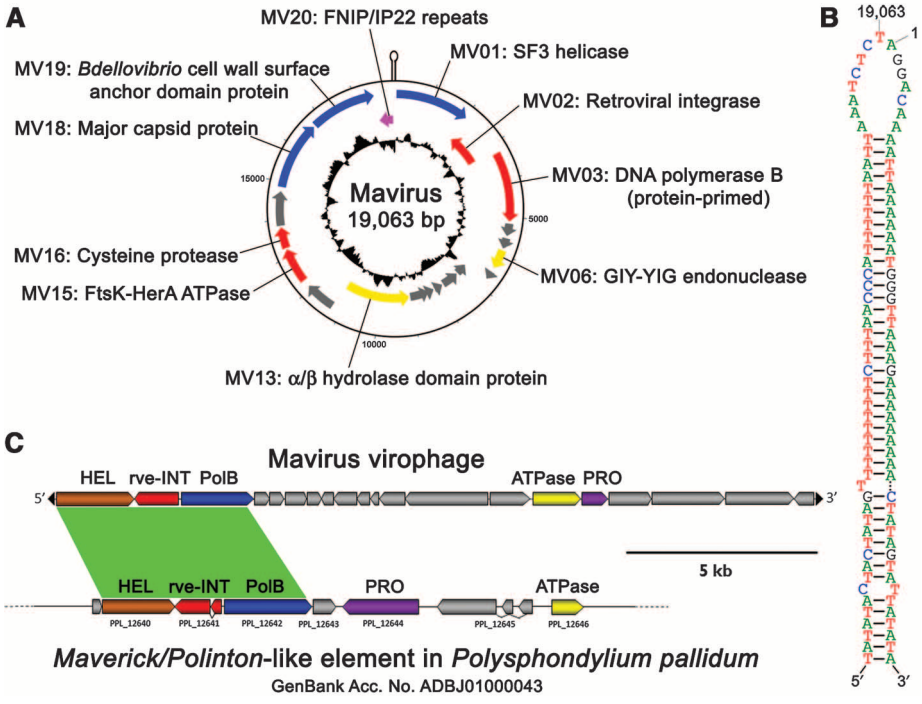

Fig. 1. (**A**) Genome diagram of the Mavirus virophage, indicating predicted protein functions. Circles from outermost to innermost represent protein-coding sequences (CDSs) on the forward strand, CDSs on the reverse strand, and G+C content (relative to the average of 30.26%). Conserved MP genes are shown in red; genes that are occasionally encoded by MP TEs are shown in blue. The FNIP/IP22 repeat-containing gene is shown in pink; additional genes with functional annotations are shaded yellow. A hairpin structure indicates the position of the inverted repeat upstream of CDS MV01. (**B**) DNA sequence and predicted secondary structure of the inverted repeat upstream of MV01. The first and last bases of the genome are numbered. (**C**) Gene organization of Mavirus (linear view of the circular genome) and of a truncated MP transposon in the slime mold *P. pallidum*. Homologous genes have the same color, and syntenic regions are shown in green.

ary hypotheses such as that MP TEs are derived from a linear plasmid (*4*), and that MP TEs and viruses of the PRD1-adenovirus lineage have evolved from a common ancestral virus (*5*, *7*). Additionally, the origin of nucleocytoplasmic large DNA viruses has been linked to MP TEs (*18*). Although one can envision several scenarios that would explain the genetic homology between Mavirus and MP TEs, two likely candidates are that Mavirus was derived from an escaped TE, or that MP TEs arose from the endogenization of an ancient relative of Mavirus. The first scenario is problematic because it requires multiple gene-capture events by MP TEs from various sources. For example, the integrase would have to be acquired from a retroelement, whereas the cysteine protease and FtsK-HerA–like ATPase would have originated in a DNA virus. In addition, the first scenario fails to explain the genetic (figs. S3 to S6) and phenotypic similarities between Mavirus and Sputnik, as well as the dependency of Mavirus on CroV. A more parsimonious explanation is that a Mavirus ancestor (MVA) gave rise to MP TEs. This may have occurred in an early eukaryote susceptible to infection by a large DNA virus (LDNAV) that coevolved with the MVA (Fig. 2). MVA replicated by parasitizing the virion factory of LDNAV and interfered with its propagation, much as occurs in the CroV-Mavirus and Mimivirus-Sputnik systems. The interference of MVA with LDNAV replication led to increased survival of the host-cell population, and selection on the host cell to stabilize the relationship with MVA. In turn, the dependence of MVA on LDNAV replication provided strong selection for MVA to be closely associated with the LDNAV host. The fortuitous acquisition of an integrase gene by MVA provided a mechanism for integration into the cellular genome, which may have conferred protection against infection by LDNAV, and selective pressure for MVA to spread throughout the host-cell population. Selection for this viral defense system in early eukaryotes and repeated endogenization of various virophages in different eukaryotic lineages, combined with accumulation of mutations, intragenomic expansion, and partial or complete loss of the provirophage in some lineages, may have led to the current diversity and widespread, albeit patchy, taxonomic distribution of MP DNA transposons (Fig. 2) (*4*, *5*).

In addition to their genetic similarities, the evolutionary relationship between Mavirus and MP TEs is supported by other observations. First, the integrases of Mavirus and MP TEs are closely related, suggesting that an ancestor of the Mavirus integrase could have catalyzed the insertion of DNA into a eukaryotic genome. We could, however, not detect endogenous forms of Mavirus in uninfected *C. roenbergensis* by PCR with Mavirus-specific primers. Second, Mavirus, as its ancestors presumably did, uses clathrin-mediated endocytosis to enter the host cell independently of its host virus, as shown by thin-section electron microscopy (fig. S1). This suggests that Mavirus can associate with the host cell even in the absence of a host virus. Third, Mavirus interferes with CroV propagation and increases the survival of the host-cell population (table S1); hence, eukaryotes that are susceptible to infection by giant viruses will gain a selective advantage if they can associate themselves with virophages. The directionality of the proposed viral transposogenesis is also supported by the shared promoter motif of Mavirus and CroV. It is hard to envision a mechanism that would have led to the de novo development from a transposon to the observed host-virus dependence, whereas it is conceivable that the promoter sequence degenerated after a MVA integrated into the host cell genome. In the light of recently described endogenous viruses (*19*), the discovery of Mavirus reveals a further facet of the intricate genetic interactions between viruses and cellular life.

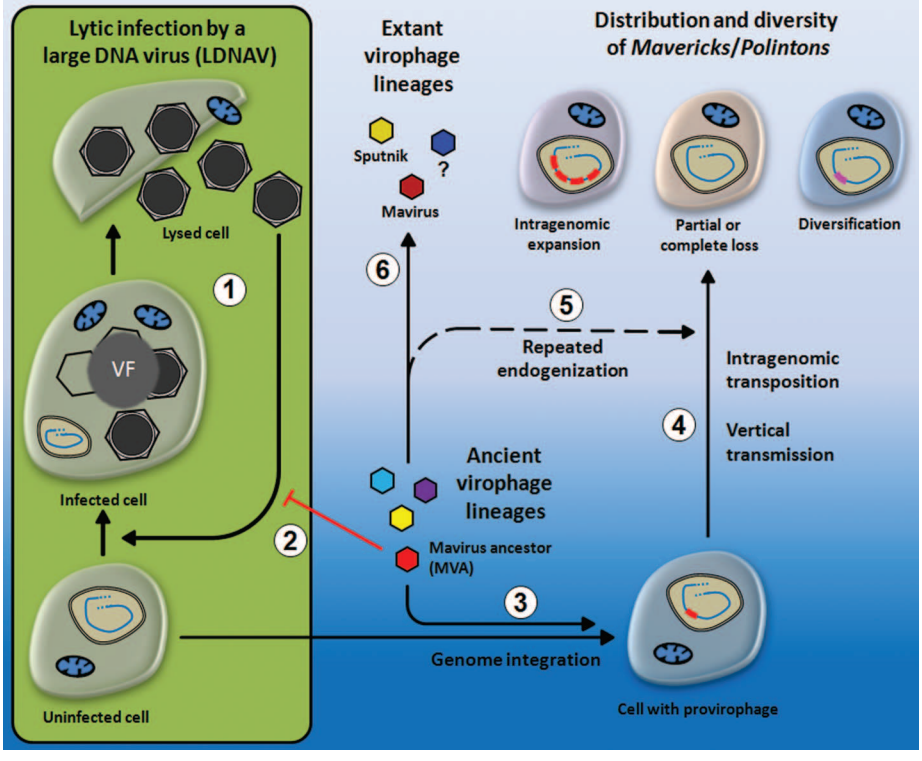

Fig. 2. Hypothesis for the origin of the *Maverick/Polinton* class of DNA transposons. (1) The infection cycle of a lytic large DNA virus (LDNAV). After virus entry, the virion factory (VF) produces progeny virions that are released during cell lysis. (2) A virophage (the Mavirus ancestor, MVA) is dependent on coinfection with the LDNAV and negatively affects its lytic infection cycle. The survival rate of the eukaryotic host population is increased by the MVA. (3) Using its integrase, the MVA inserts its genome into the host cell chromosome. (4) Over millions of years, the provirophage is vertically transmitted, accumulates mutations, expands horizontally in some eukaryotic lineages [e.g., *Trichomonas vaginalis* (*5*)], or suffers partial or complete loss in other lineages [e.g., Ascomycota (*4*)]. (5) These evolutionary processes are complemented by repeated endogenization of Mavirus-like virophages and result in the observed distribution and diversity of MP TEs. (6) Of presumably multiple ancient lineages of virophages, only the one from which Mavirus descended gave rise to the MP TEs. The Sputnik virophage has probably evolved from a different lineage, and many others may remain to be discovered.

References and Notes

1. G. Bourque, *Curr. Opin. Genet. Dev.* **19**, 607 (2009).
2. C. Feschotte, E. J. Pritham, *Annu. Rev. Genet.* **41**, 331 (2007).
3. C. Feschotte, E. J. Pritham, *Trends Genet.* **21**, 551 (2005).
4. V. V. Kapitonov, J. Jurka, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4540 (2006).
5. E. J. Pritham, T. Putliwala, C. Feschotte, *Gene* **390**, 3 (2007).
6. E. J. Pritham, *J. Hered.* **100**, 648 (2009).
7. M. Krupovič, D. H. Bamford, *Nat. Rev. Microbiol.* **6**, 941 (2008).
8. S. D. Benson, J. K. Bamford, D. H. Bamford, R. M. Burnett, *Cell* **98**, 825 (1999).
9. M. G. Fischer, M. J. Allen, W. H. Wilson, C. A. Suttle, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19508 (2010).
10. Materials and methods are available as supporting material on *Science* Online.
11. B. La Scola *et al.*, *Nature* **455**, 100 (2008).
12. J.-M. Claverie, C. Abergel, *Annu. Rev. Genet.* **43**, 49 (2009).
13. M. Legendre *et al.*, *Genome Res.* **20**, 664 (2010).
14. E. V. Koonin, S. Zhou, J. C. Lucchesi, *Nucleic Acids Res.* **23**, 4229 (1995).
15. X. Gao, D. F. Voytas, *Trends Genet.* **21**, 133 (2005).
16. W. Bao, V. V. Kapitonov, J. Jurka, *Mob. DNA* **1**, 3 (2010).
17. N. J. Strömsten, D. H. Bamford, J. K. H. Bamford, *J. Mol. Biol.* **348**, 617 (2005).
18. J. Filée, N. Pouget, M. Chandler, *BMC Evol. Biol.* **8**, 320 (2008).
19. W. E. Johnson, *PLoS Genet.* **6**, e1001210 (2010).

20. We thank B.N. Ross and G.D. Martens of the UBC Bioimaging Facility and A.M. Chan for assistance with

electron microscopy; E. Zaikova for qPCR assistance; and E. J. Pritham, J. Davies, and E. Nomis for constructive comments. Supported by the Natural Science and Engineering Research Council of Canada (NSERC) Discovery Grants Program, the Tula Foundation through the Centre for Microbial Diversity and Evolution, and fellowships awarded to M.G.F. by the Gottlieb Daimler- and Karl Benz-Foundation, Germany, and the University of British Columbia. The genome sequence of Mavirus has been deposited in GenBank under the accession number HQ712116.

Supporting Online Material
www.sciencemag.org/cgi/content/full/science.1199412/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S3
References

22 October 2010; accepted 8 February 2011
Published online 3 March 2011;
10.1126/science.1199412

A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis

Gira Bhabha,[1] Jeeyeon Lee,[2]* Damian C. Ekiert,[1] Jongsik Gam,[2] Ian A. Wilson,[1] H. Jane Dyson,[1] Stephen J. Benkovic,[2] Peter E. Wright[1]†

Conformational dynamics play a key role in enzyme catalysis. Although protein motions have clear implications for ligand flux, a role for dynamics in the chemical step of enzyme catalysis has not been clearly established. We generated a mutant of *Escherichia coli* dihydrofolate reductase that abrogates millisecond-time-scale fluctuations in the enzyme active site without perturbing its structural and electrostatic preorganization. This dynamic knockout severely impairs hydride transfer. Thus, we have found a link between conformational fluctuations on the millisecond time scale and the chemical step of an enzymatic reaction, with broad implications for our understanding of enzyme mechanisms and for design of novel protein catalysts.

Protein motions are critical for biological functions, but their precise role in enzyme catalysis remains unclear (*1*, *2*). Although there is convincing evidence that conformational fluctuations are essential for the mediation of substrate and cofactor binding as well as product release and can be rate-limiting for enzyme turnover (*3*–*6*), the importance of protein flexibility for progression along the chemical reaction coordinate remains a matter of debate (*7*, *8*). In one view, electrostatic preorganization of the active site is considered to account fully for enzyme catalysis (*9*); however, hydrogen tunneling experiments suggest that a static model is inadequate and that fluctuations that reorganize the active site are required for the chemical step (*7*, *10*, *11*), which involves conformational sampling to facilitate hydrogen transfer as well as the transfer itself.

Escherichia coli dihydrofolate reductase (*ec*DHFR) is a paradigm for understanding the relationship between structure, dynamics, and catalysis (*4*, *12*–*14*). DHFR catalyzes the stereospecific reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) using reduced nicotinamide adenine dinucleotide phosphate (NADPH) as cofactor. Five intermediates have been identified in the steady-state catalytic cycle (*7*): E:NADPH, E:NADPH:DHF, E:NADP$^+$:THF, E:THF, and E:NADPH:THF (Fig. 1). Large conformational changes are observed in the Met20 loop (residues 9 to 24) between the Michaelis complex E:NADPH:DHF [modeled by E:NADP$^+$:folate (FOL) for structural studies] and the product complex, E:NADP$^+$:THF (*13*). The Met20 loop adopts two dominant conformations: the "closed" conformation, in which the loop packs tightly against the nicotinamide ring of the cofactor, and the "occluded" conformation, in which the loop projects into the active site and sterically blocks the nicotinamide-binding pocket. The holoenzyme (E:NADPH) and the model Michaelis complex (E:NADP$^+$:FOL) are in the closed conformation, whereas the three product complexes (E:NADP$^+$:THF, E:THF, and E:NADPH:THF) adopt the occluded conformation (Fig. 1), which is stabilized by hydrogen bonds between Asn23 and the backbone and side chain of Ser148 (*13*, *15*).

In crystal structures of human and other vertebrate DHFRs, the Met20 loop is invariably closed (*16*). Alignment of the human and *E. coli* DHFR protein sequences suggests that the occluded conformation is destabilized in the human enzyme because of an inability to form the stabilizing hydrogen bond made by the S148 side chain in *ec*DHFR (fig. S1). In addition, a

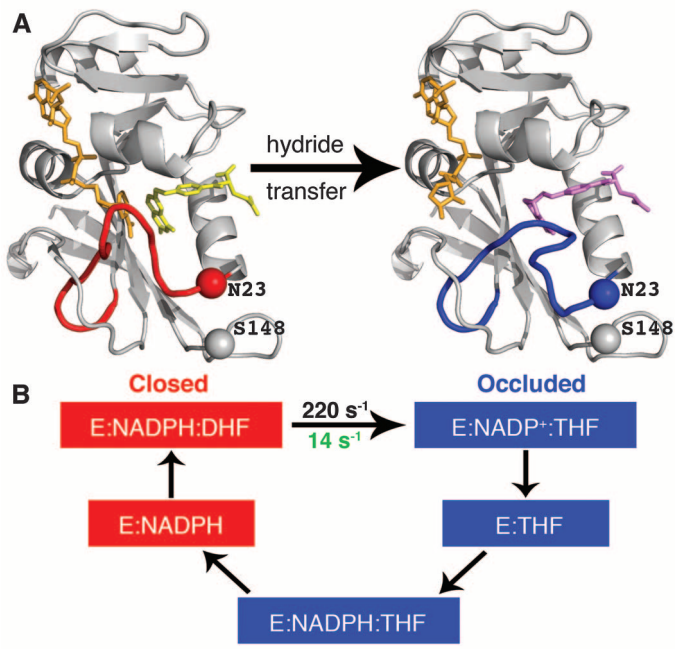

Fig. 1. Conformational changes that occur during the *E. coli* DHFR catalytic cycle. (**A**) (Left) Illustration of E:NADP$^+$:FOL crystal structure (1RX2, model for the Michaelis complex, E:NADPH:DHF) in the closed conformation. (Right) Crystal structure of E:NADP$^+$:ddTHF (1RX4, model for the product complex) in the occluded conformation. NADP$^+$ is shown in orange; FOL is shown in yellow, and ddTHF is shown in purple. Red indicates Met20 loop in the closed conformation; blue indicates Met20 loop in the occluded conformation. The sites of mutation, N23 and S148, are shown as spheres. (**B**) Intermediates in the wild-type *E. coli* DHFR catalytic cycle. Intermediates shown in red are in the closed conformation, and those in blue are in the occluded conformation. Before hydride transfer, the Met20 loop is in the closed conformation, in which it packs tightly against the nicotinamide ring of NADP$^+$. After hydride transfer, the Met20 loop adopts the occluded conformation, in which the nicotinamide ring of NADP$^+$ is sterically hindered from binding in the active site. NADP$^+$ undergoes a concurrent conformational change in which the nicotinamide ring is expelled from the binding pocket, initiating NADP$^+$ release from the ternary product complex. The rate of hydride transfer in the wild-type and N23PP/S148A mutant enzyme is indicated in black and green, respectively. The mutation alters the pathway used for product and NADP$^+$ release, as shown in fig. S2.

[1]Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. [2]Department of Chemistry, Pennsylvania State University, University Park, PA 16802, USA.

*Present address: College of Pharmacy, Ajou University, San 5, Woncheon-dong, Yeongtong-Gu, Suwon 443-749, Korea.
†To whom correspondence should be addressed. E-mail: wright@scripps.edu